

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Dynamic Process Model Parameter Estimation by Global System Analysis

Shigeru Kashiwaya

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.74635>

Abstract

Global system analysis (GSA) was applied to parameter estimation of dynamic process models. First, the posterior distribution of the model parameters was estimated by quasi-Monte Carlo (QMC) simulations or uncertainty analysis. The expected variance of the estimated parameters by GSA was in general smaller than those were obtained by local search for the maximum likelihood. Second, sensitivity analysis was performed as an alternative application of GSA for the same mathematical models and testing data. The total effect index should serve as a quantitative measure of the robustness of each estimated parameter. Two process models were studied to demonstrate effectiveness of the proposed methodology based on GSA: a bio-reactor and a catalytic reactor. Parallelised computation allowed for sampling as many as 500,000 combinations of the model parameters in reasonable amount of time.

Keywords: parameter estimation, global system analysis, quasi-Monte Carlo, sensitivity analysis, high-performance computation

1. Introduction

We try to interpret, understand, and extract knowledge by analysing the data obtained from measurements, experiments and recording the behaviour of various processes that we encounter in the industries and real life.

Parameter estimation of dynamic process models is investigated in this study. To validate a mathematical process model, we need to match the model prediction with the measurement or the experiment data.

We assume that we can construct a certain mathematical or a first principle model to represent a chemical or biochemical process by a set of algebraic (either linear or nonlinear) and differential (either ordinary or partial) equations, that is, differential algebraic equations (DAEs).

The Bayesian statistics provides an ideal framework for the estimation of the mathematical parameters of predictive models, which is inherently *stochastic* in that all models make imperfect predictions because of the process variance, observation errors, model selection uncertainties, and so on.

The uncertainty of the estimated parameters must be evaluated in terms of *accuracy* or *posterior* distribution in the Bayesian statistics context.

Although maximum likelihood estimation (MLE) has been the standard tool for solving the parameter estimation problems, some shortcomings have been recognised in terms of its properties and the *conventional* solution algorithms for the MLE (see [1]):

1. Gradient-based methods (GBM) may miss the global optimum or *true* MLE because of the nature of such solvers that are based on local search mechanisms.
2. The confidence region for the parameters can be very large when the curvature of the likelihood function at an optimal point is extrapolated without knowing the global distribution of the likelihood function around the point estimate.
3. If the model equations contain singularities or structural redundancy, correlations among the estimated parameter may be observed.

Global system analysis (GSA) is designed to provide advanced features compared with the *conventional* analytical methods such as GBM and MLE. GSA is not precisely the same theoretical framework as Bayesian method; however, we can apply GSA as a version of computational Bayesian methodology to estimate model parameters.

First, GSA is a global search procedure as opposed to GBM such as Newton's method where we must start the search with some *initial guess* so that the solution can be trapped by a local optimal point that is not necessarily the global optimum.

Second, the Bayesian Cramer-Rao (BCR) bounds offer tighter lower bounds than the conventional Cramer-Rao lower bound that is based on Fisher information. GSA can calculate BCR with assumption of a prior that can be uniform (non-informative) or any probability distribution such as Gaussian.

Third, the posterior distribution of the estimated model parameters obtained by GSA correctly reflects the non-linearity of the process. Some process kinetics equations such as Arrhenius equation present structural correlation, for example, between pre-exponent factors and activation energies. Even in such a case, GSA can correctly capture the interdependency among the parameters so that the posterior distribution of the parameters accuracy is evaluated properly.

Since GSA algorithm is deterministic, one can decide the required number of samples prior to numerical simulation depending on the required accuracy or convergence. The computation can also be parallelised so that the actual CPU time can be significantly accelerated. This is a

clear advantage over Markov Chain Monte Carlo (MCMC) sampling where we need to monitor the convergence to make sure the sample distribution approached the intended convergence criterion.

GSA is applied to parameter estimation taking the likelihood function as the *response* to the *factors*, which are in fact model parameters that are considered stochastic variables in the Bayesian statistics framework. The parameter search space may be either non-informative or informative in GSA: low discrepancy sequence (LDS) such as Sobol sequence is considered non-informative; or normal (Gaussian) distribution may also be assumed to perform sampling in the multidimensional parameter space (more distribution functions such as gamma or inverse-gamma will be available in the GSA tool that was used for this study).

Posterior of the model parameters can be calculated simply by integrating the sampled likelihood (*response*) over the entire parameter space (*factors*). We can then calculate posterior mean and variance as well as maximum a posterior (MAP) estimate.

Sensitivity analysis can also be performed by sampling in the hyper-dimensional space just like in the case of uncertainty analysis. The factor sensitivity table gives information about the influence of each parameter on the response, which is the likelihood function of the model postulated for the investigation.

If the process model contains some singularity or correlations among the parameters, the parameter estimation work gets very difficult, if not theoretically impossible. Although a number of different procedures have been proposed for re-parameterisation of the process models, since such *manipulations* are not essential in the physical significance of the processes of interest, the estimated parameters that are reformulated would not be very useful to understand the fundamental principles of the actual processes.

Since GSA is computational resource intensive, parallelisation should be considered to make the analysis tractable. Recent advance in the software technology made it possible to run relatively large GSA with as many as 500,000 samples executed even on a common laptop PC with a multi-core CPU

2. Parameter estimation by global system analysis

GSA helps to investigate the global behaviour of any system models. In contrast to simulation, in which specific values are assigned to system inputs and the values of key outputs are reported, in GSA, it is possible to specify a range of input values of interest and therefore obtain a range of outputs.

GSA can perform three types of analysis:

- Scenario or parametric
- Uncertainty
- Sensitivity

The purpose of each type of analysis differs widely as follows:

- Scenario or parametric sensitivity analysis is a series of simulations where the values of one or more factors are varied over a grid to investigate the resulting changes in one or more responses. It allows an assessment of the impact of different factors on each response.
- Uncertainty analysis propagates the uncertainty in the factors to the responses. The values returned represent the actual uncertainty associated with the responses.
- Sensitivity analysis provides metrics which indicate how factors and their uncertainty influence responses. However, the values obtained do not represent actual sensitivities.

When the likelihood function is taken as the objective or the only output from the system whereas the model parameters are given as inputs, GSA can be utilised as a global search and uncertainty evaluation for the model parameter estimation. The uncertainty analysis was used to identify the parameter point estimates and confidence intervals. In addition, the sensitivity analysis was performed to quantify the influence of each parameter on the output or prediction of the process models that are investigated.

2.1. Parameter estimation

Bayesian approach to the model parameter estimation can be presented as Eq. 1.

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto L(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (1)$$

where $p(\boldsymbol{\theta})$ designates a prior distribution of the model parameters, $L(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood with which one observes response $\mathbf{y}(\hat{y}_{ijk})$, and $\pi(\boldsymbol{\theta}|\mathbf{y})$ is the posterior.

When certain a priori knowledge is available about the process, it can be incorporated in the Bayesian framework as a prior. In this study, however, we assumed a flat or homogeneous prior over the parameter search space.

The MAP estimate is defined by the parameter that maximises the posterior probability given as Eq. 2.

$$\boldsymbol{\theta}_{MAP} = \arg \max_{\boldsymbol{\theta}} (\pi(\boldsymbol{\theta}|\mathbf{y})) \quad (2)$$

The system responses are parameterised by a set of parameters in the following equation, where $\boldsymbol{\theta}^*$ denotes *true* parameters that characterise the system. We assume additive errors that are *i.i.d*, that is, *independent and identically distributed following a normal distribution*.

$$\hat{y}_{ijk} = \mathbf{g}_j(\mathbf{x}_i; \boldsymbol{\theta}^*) + \varepsilon_{ijk} \quad (i = 1, 2, \dots, NE) \quad (3)$$

The probability f with which one observes measurement \hat{y}_{ijk} is formulated as Eq. 4.

$$f(\hat{y}_{ijk}|\theta, \sigma_{ijk}^2) = \left(2\pi\sigma_{ijk}^2\right)^{-1/2} \exp\left(-\frac{(\hat{y}_{ijk} - \mathbf{g}_j(\mathbf{x}_i; \theta))^2}{2\sigma_{ijk}^2}\right) \quad (4)$$

The log-likelihood $LL(\theta)$ is defined as the logarithm of likelihood $L(\theta) \equiv L(\mathbf{y}|\theta) = \prod_{ijk} f(\hat{y}_{ijk}|\theta, \sigma_{ijk}^2)$ that is a function of parameters $\theta(\theta_1, \dots, \theta_q)$.

$$\begin{aligned} LL(\theta) &\equiv -\log L(\theta) \\ &= -\log \left[\prod_{i=1}^{NE} \prod_{j=1}^{NV_i} \prod_{k=1}^{NM_{ij}} f(\hat{y}_{ijk}|\theta, \sigma_{ijk}^2) \right] = \frac{n}{2} \log(2\pi) \\ &\quad + \frac{1}{2} \sum_{i=1}^{NE} \sum_{j=1}^{NV_i} \sum_{k=1}^{NM_{ij}} \left[\log(\sigma_{ijk}^2) + \frac{(\hat{y}_{ijk} - \mathbf{g}_j(\mathbf{x}_i; \theta))^2}{\sigma_{ijk}^2} \right] \end{aligned} \quad (5)$$

where $f(\hat{y}_{ijk}|\theta, \sigma_{ijk}^2)$, likelihood of measurement \hat{y}_{ijk} in normal (Gaussian) error with mean θ and variance σ_{ijk}^2 ; $\mathbf{g}_j(\mathbf{x}_i)$, simulated model response of variable j in experiment i ; $LL(\theta) \equiv -\log L(\mathbf{y}|\theta)$, logarithmic likelihood function; n , number of measurements taken during all experiments; NE , number of experiments performed; NM_{ij} , number of measurements of the j -th variable in experiment i ; NV_i , number of variables measured in the i -th experiment; q , number of model parameters; \mathbf{x}_i , operating conditions of the i -th experiment; \hat{y}_{ijk} , k -th measured value of variable j in experiment i ; σ_{ijk}^2 , variance of the k -th measurement of variable j in experiment i ; $\theta = \theta(\theta_1, \dots, \theta_q)$, model parameters to be estimated; and σ_{ijk}^2 , variance of the k -th measurement of variable j in experiment i .

The MLE is defined by the parameter that minimises the log-likelihood function:

$$\theta_{MLE} = \arg \min_{\theta} (LL(\theta)) \quad (6)$$

2.2. Global system analysis

GSA is based on drawing samples from the system under study; that is, taking different values for each factor and calculating the corresponding response values. One of the features and basic principles underlying GSA is that extracting random samples from a system permits the estimation of its mean value

Considering the atypical surface shown in **Figure 1** below on the left, **Figure 2** on the right indicates what its average value will be for a number of random sample points. The dispersion of the values is due to the fact that this process is repeated 100 times for each number of samples, each time generating a different random set of sample points. This is indicative of how accurate the results will be for different numbers of samples

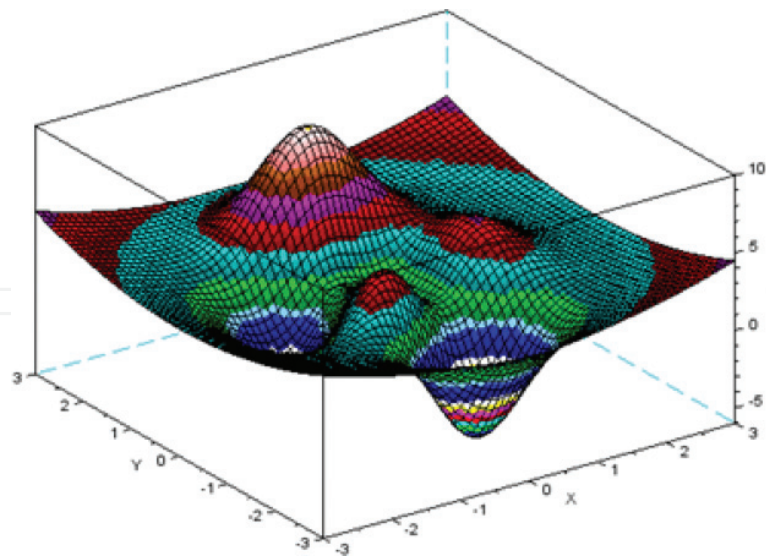


Figure 1. Response surface for two inputs.

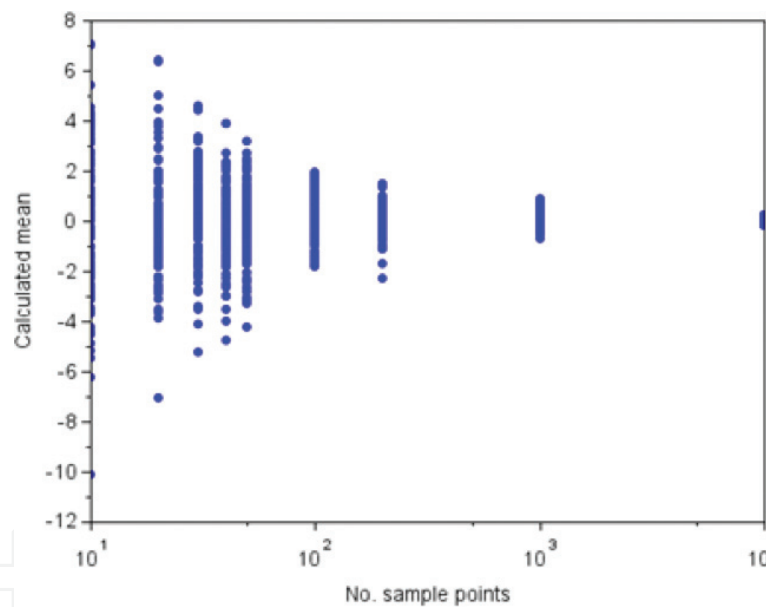


Figure 2. Spread of averages for different number of samples.

It is clear that by increasing the number of points, values gradually become closer to the true mean value of the surface. Increasing the number of sample points will generally lead to more accurate results.

2.3. Uncertainty analysis

An uncertainty analysis can be used to determine what effect the uncertainty in the factors has on the uncertainty in the responses. This type of analysis can be accomplished using the Monte

Carlo method. The Monte Carlo method performs multiple model evaluations (simulations) with deterministic and/or probabilistic factors. The results of these evaluations are then used to determine the uncertainty in the responses. This method is appropriate for any type of model, regardless of complexity.

Pseudo-random sampling generates a more realistic random sample, that is, resembling a sample drawn from a distribution. This method should be used to represent realistic sampling, for example, a clinical trial in which 200 people are sampled from the general population.

Quasi-random (Sobol) sampling (**Figure 3**) allows for better coverage of space compared to pseudo-random sampling (**Figure 4**). Sobol sampling is recommended unless a more realistic random behaviour is required.

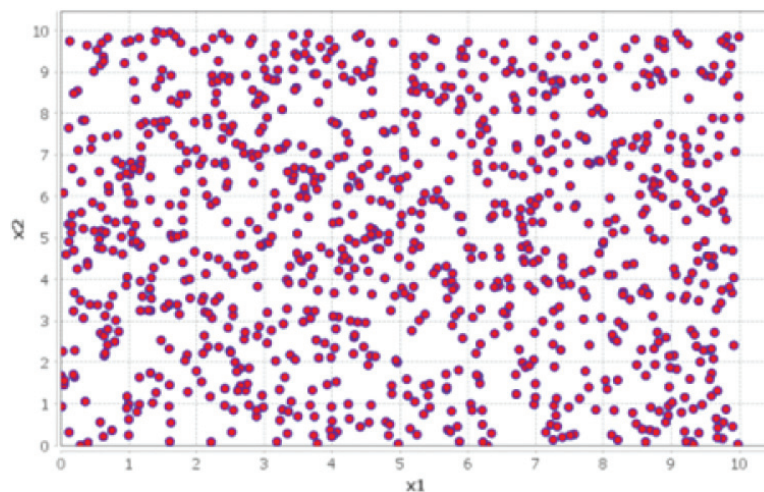


Figure 3. Quasi-random (Sobol) sampling.

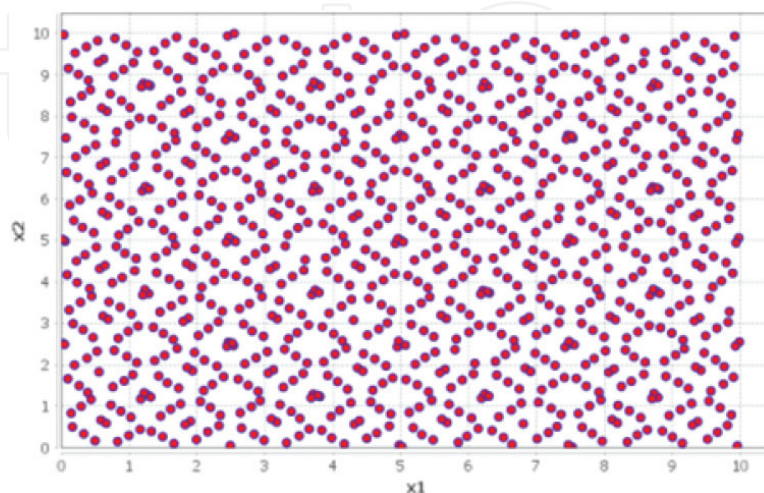


Figure 4. Pseudo-random sampling.

2.4. Sensitivity analysis

Sensitivity analysis can be used to determine which factor contributes the most to the uncertainty in a response. In GSA, the calculated sensitivity measures are sensitivity indices rather than mathematical sensitivities (derivatives). For example, consider the following equation

$$y = x_1^2 + 3x_2 \quad (7)$$

where all variables are algebraic and both x_1 and x_2 are assumed to have values between 0 and 10. Their mathematical sensitivities are given by:

$$\frac{\partial y}{\partial x_1} = 2x_1 \quad (8)$$

$$\frac{\partial y}{\partial x_2} = 3 \quad (9)$$

where y , x_i , algebraic variables; and $\frac{\partial y}{\partial x_i}$, partial derivative of y with respect to x_i .

These mathematical sensitivities are informative at the point where they are computed, but are limited when the factors are uncertain, and the model is of unknown linearity, as they do not provide for an exploration of the rest of the input factor space. In contrast, GSA calculates sensitivity measures that are averages over the entire space of interest. Two different indices are available: elementary effects and variance-based indices, which provide relative metrics.

In both cases, indices for the *first order* and *total effect* are calculated. According to Sobol [2, 3], any function expressed either as follows:

$$y = f(u_1, \dots, u_{n_f}) \quad (10)$$

where f , function; n_f , number of factors; u_i , factor i ; and y , response.

or as a combination of different functions:

$$y = f_0 + \sum_{i=1}^{n_f} f_i(u_i) + \sum_{i < j}^{n_f} f_{ij}(u_i, u_j) + \dots + f_{1,2,\dots,n_f}(u_1, u_2, \dots, u_{n_f}) \quad (11)$$

where f_i , first-order function; f_{ij} , second-order function; and $f_{1,2,\dots,n_f}$, n_f th order function.

The *first-order effect* of factor u_i is f_i . In contrast, the *total effect* is the sum of all functions of all orders where u_i appears as an argument

Elementary effects are an efficient method for larger, more complex models with a large number of factors. This method requires a limited number of samples. However, it is not as accurate as the variance-based method and should only be used to identify the most important factors.

For elementary effects, the sensitivity indices are calculated based on the following approximation.

$$EE_{i,j} = \frac{y_j(u_1, \dots, u_i + \Delta, u_n) - y_j(u_1, \dots, u_i, u_n)}{\Delta} \quad \forall i \in \{1, \dots, n\} \quad (12)$$

where $EE_{i,j}$, elementary effect for factor i for sample j ; u_i , factor i ; y_j , response of sample j ; and Δ , difference.

The mean values are calculated as follows:

$$\mu_i = \frac{1}{n} \sum_{j=1}^n EE_{i,j} \quad (13)$$

where n , number of samples; and μ_i , mean elementary effect for factor i .

Although this appears to be an approximation of the mean partial derivative, it is typically not an accurate estimate of the actual mean sensitivity. Among other differences, the elementary effect also considers the variance in each factor. For example, considering the following equation $y = x_1 + x_2$, the value of the mathematical sensitivity for both x_1 and x_2 is one, but if the variance in x_1 is larger than x_2 , then $\mu_1 > \mu_2$.

The elementary effects method also calculates the standard deviation as follows.

$$\sigma_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (EE_{i,j} - \mu_i)^2} \quad (14)$$

where σ_i is the standard deviation of elementary effect for factor i .

The advantage of elementary effects analysis is that it is a numerically efficient way to screen many factors at the expense of accuracy. It has been shown in practice that the relative value of the sensitivity is fairly accurate, so values sorted from high to low do typically follow the order of the most influential to the least influential factor.

Variance-based methods are more accurate but more computationally demanding and require (many) more samples than the elementary effects method. The variance-based method was applied to the following numerical examples.

The metrics used are based on variances and are indicative of what part of the variance, in the response, is attributed to the variance in each factor. Note that the relative contribution of the factors to the variance in the response can be sorted differently for each response.

This method distinguishes between first order and total effect indices. Generally, one can split any function.

Variance-based sensitivity indices measure the influence of individual factors on responses. The variance-based sensitivity analysis method is based on the Saltelli's method [4], and the formulae used to estimate the sensitivity indices are those proposed in [5]. There are two types of variance-based sensitivity indices (see [2, 3])

- The first-order effect, that is, the direct effect of factor i on response y :

$$S_i = \frac{\sigma_{i,y}^2}{\sigma_y^2} \quad (15)$$

where S_i , first-order effect index for factor i ; $\sigma_{i,y}^2$, variance attributed to factor i on response y ; and σ_y^2 , variance observed for response y .

- The total effect, that is, the total effect of factor i on response, y :

$$S_{i,T} = \frac{\sigma_{i,T,y}^2}{\sigma_y^2} \quad (16)$$

where $S_{i,T}$, total effect index for factor i ; and $\sigma_{i,T,y}^2$, variance attributed to factor i on response y through all factors.

In this case, the influence of factor i through other variables is taken into consideration.

In variance-based sensitivity analysis, the sensitivity indices are based on variances. Note that the relationships below can be inferred from the definitions of the factors:

$$\sum_{i \in \text{factors}} S_{j,i} \leq 1 \quad (17)$$

$$S_{j,i,T} \geq S_{j,i} \quad (18)$$

$$\sum_{i \in \text{factors}} S_{j,i,T} \geq 1 \quad (19)$$

where $S_{j,i}$, first-order effect index for factor i in response j ; and $S_{j,i,T}$, total effect index for factor i in response j .

It is clear that mathematical sensitivities and sensitivity indices have very different meanings. For simplicity, consider the sensitivity index as a normalised metric which indicates what part of the variance in the response can be attributed to which factor.

3. Numerical case studies

gPROMS[®], a general-purpose DAE solver for process modelling, was used to compute the likelihood function values based on the measurement data with some artificial errors for the demonstration purposes. It was also used to compute the MLE (by GBM) to set a reference for comparison with the results by the proposed method, GSA-based inference.

3.1. Case study 'A': bioreactor

A benchmark problem in genome research in [6] is reproduced. They applied bootstrapping to minimise the expected variance of the estimated parameters.

3.1.1. Problem statement

In silico experimental data for an organism in a chemostat as shown in **Figure 5** is presented. A mathematical model is set up based on a fictive network structure.

After reaching a steady state, the flow rates q_{in} , q_{out} as well as the concentration of the substrate in the feed c_{in} are changed. Measurements are available for three metabolites, M_1 , M_2 , and M_3 that function as an enzyme E, representing a small biochemical network of the organism, and for biomass B and substrate S (**Table 1**).

Those parameters as listed in **Table 2** are assumed known prior to the experiment.

Since number of such experiments that can be performed is rather limited in reality, GSA or a Bayesian approach is anticipated to perform better than MLE or minimum mean square error (MMSE) methods in that the latter tends to provide overfitting of the parameters in terms of

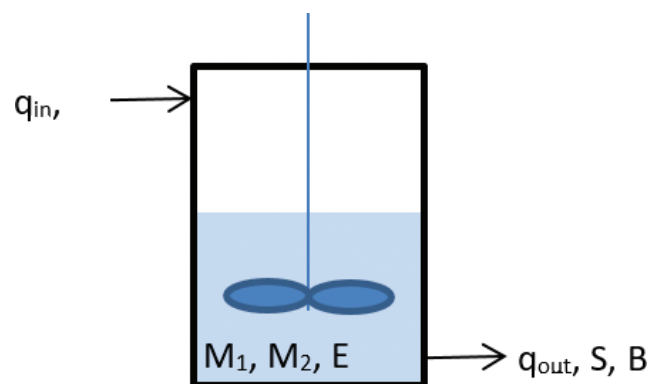


Figure 5. Bioreactor.

'Known' kinetics	Parameter	Value	Unit
Synthesis of M_1 values	r_{1max}	2.4×10^4	$\mu\text{mol/g/h}$
	K_S	0.4437	$\mu\text{mol/g}$
Affinity M_1 – E value	K_{M1}	12.2	$\mu\text{mol/g}$
Degradation of M_2 values	r_{3max}	3.0×10^6	$\mu\text{mol/g/h}$
	K_{M2}	10	$\mu\text{mol/g}$

Table 1. Additional information – 'known' parameters.

Time [hour]	c_{in} [g/litre]	q_{in} , q_{out} [litre/h]
0–20	2	0.25
20–30	2	0.35
30–60	0.5	0.35

Table 2. Experimental procedure.

the observation or measured data, whereas the former should calculate more robust estimate of parameters that are integrated over the posterior distribution.

$$\dot{V} = q_{in} - q_{out} \quad (20)$$

$$\dot{B} = \left(\mu - \frac{q_{in}}{V} \right) B \quad (21)$$

$$\dot{S} = \frac{q_{in}}{V} (c_{in} - S) - r_1 Mw B \quad (22)$$

$$r_1 = r_{1max} \frac{S}{K_S + S} \quad (23)$$

$$r_2 = k_2 E \frac{M_1}{K_{M1} + M_1} \quad (24)$$

$$r_3 = r_{3max} \frac{M_2}{K_{M2} + M_2} \quad (25)$$

$$r_{syn} = k_{synmax} \frac{K_{IB}}{K_{IB} + M_2} \quad (26)$$

$$\dot{M}_1 = r_1 - r_2 - \mu M_1 \quad (27)$$

$$\dot{M}_2 = r_2 - r_3 - \mu M_2 \quad (28)$$

$$\dot{E} = r_{syn} - \mu E \quad (29)$$

$$\mu = Y_{X/S} r_1 \quad (30)$$

where B , biomass concentration, (g/litre); c_{in} , substrate concentration in feed, (g/litre); E , enzyme ($=M_3$) concentration, ($\mu\text{mol/g}$); k_2 , turnover number to produce M_2 , (litre/h); K_{IB} , inhibition of enzyme synthesis by M_2 , ($\mu\text{mol/g}$); K_{M1} , Michaelis constant - M_1 uptake, ($\mu\text{mol/g}$); K_{M2} , Michaelis constant - M_2 uptake, ($\mu\text{mol/g}$); K_S , Michaelis constant - substrate uptake, ($\mu\text{mol/g}$); k_{synmax} , maximum enzyme synthesis rate, ($\mu\text{mol/g/h}$); M_i , metabolite i ($i = 1, 2, 3$) concentration, ($\mu\text{mol/g}$); Mw , molar mass of substrate (243.3×10^{-6}), (g/ μmol); q_{in} , feed flow rate, (litre/h); q_{out} , product flow rate, (litre/h); r_{1max} , maximum rate constant-reaction 1, ($\mu\text{mol/g/h}$); r_{3max} , maximum rate constant - reaction 3, ($\mu\text{mol/g/h}$); r_i , uptake rate to produce M_i ($i = 1, 2, 3$), ($\mu\text{mol/g/h}$); r_{syn} , enzyme synthesis rate, ($\mu\text{mol/g/h}$); S , substrate concentration, (g/litre); V , reactor holdup volume, (litre); $Y_{X/S}$, biomass yield coefficient, (g/ μmol); and μ , biomass specific growth rate, (h^{-1}).

The feed flow rate and the substrate concentration in feed are changed as shown in **Table 3** and **Figure 6**.

3.1.2. Numerical study

GSA-based parameter estimation was attempted as well as the gradient-based method or local search as the benchmark. In **Figures 7–11**, each dot in blue shows an individual measurement

	$Y_{X/S}$	k_2	k_{synmax}	K_{IB}
<i>Maximum likelihood estimators</i>				
Likelihood =	-485			
MLE	7.00E-05	6.43E + 06	6.80E-03	0.2
Std. Dev.	6.24E-07	1.85E + 05	5.70E-04	0.2
<i>Empirical Bayes estimators (GSA)</i>				
Likelihood =	-476			
MAP	6.91E-05	6.70E + 06	6.26E-03	2.7
Posterior Mean	6.94E-05	6.65E + 06	6.31E-03	5.3
Posterior Variance	5.70E-07	1.21E + 05	1.30E-04	4.8
<i>Factor sensitivity table</i>				
First order effect	0.021	0.02	0.054	0.003
Total effect	0.768	0.697	0.906	0.006

Table 3. Case study 'A' estimated parameters.

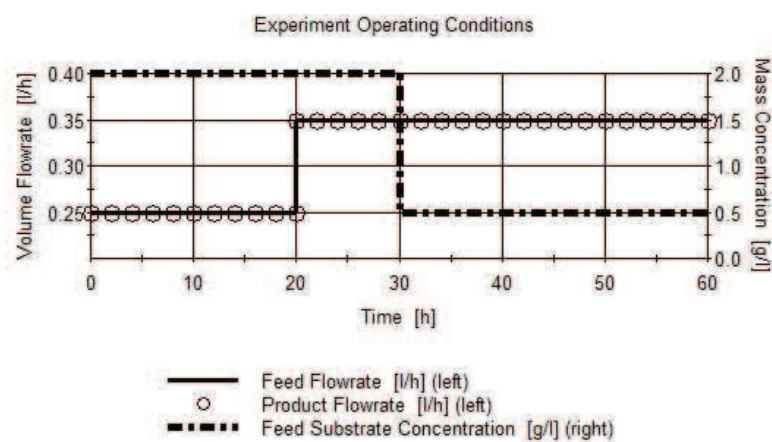


Figure 6. Operating conditions of the experiments.

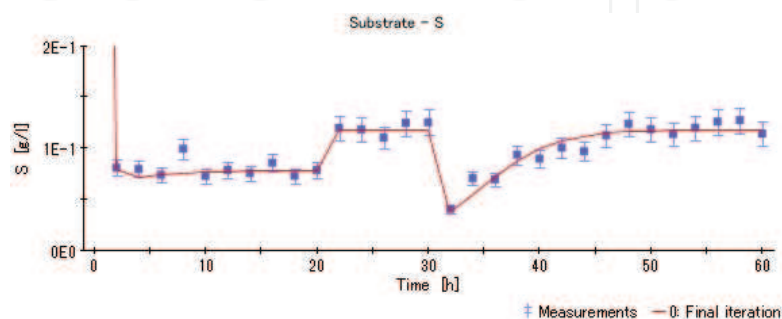


Figure 7. Overlay plot for trajectory of substrate S.

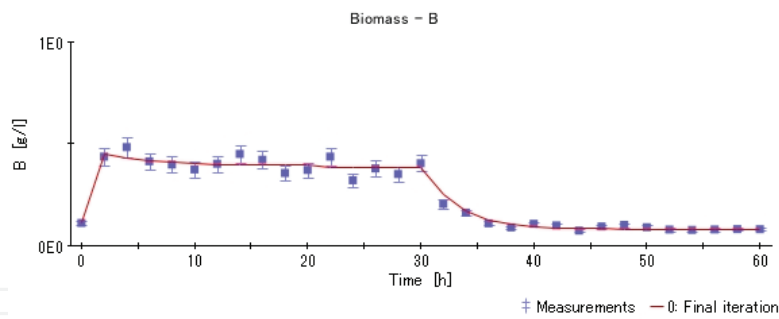


Figure 8. Overlay plot for trajectory of biomass B.

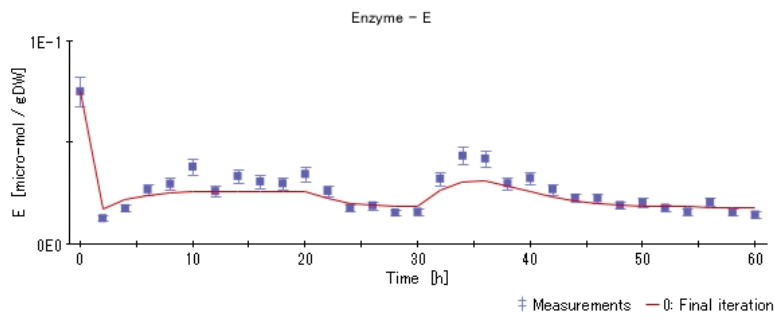


Figure 9. Overlay plot for trajectory of enzyme E (M_3).

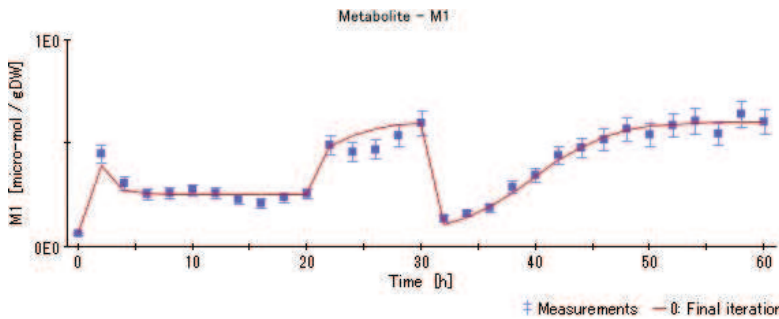


Figure 10. Overlay plot for trajectory of metabolite M_1 .

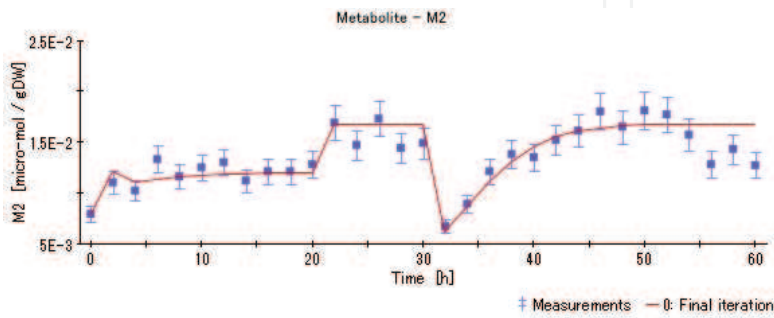


Figure 11. Overlay plot for trajectory of metabolite M_2 .

in dynamic experiments. The error bar on each dot signifies standard deviation of the measurement data (linear variance model). The solid lines in red are MLE estimation by the GBM solver (gPROMS®).

A constant linear variance model is assumed for each point of the measurement data where $\omega = 0.02$.

$$\sigma^2 = \omega^2 z^2 \quad (31)$$

GSA sampling was performed by Sobol sequence involving four parameters $Y_{X/S}$, k_2 , k_{synmax} , K_{IB} . A total of 500,000 sample points were generated in terms of combinations of these four parameters. Marginal posterior for each parameter is plotted in **Figures 12–15**.

Figures 16–21 are 3D presentation of the posterior in terms of the likelihood function.

Both $Y_{X/S}$ and k_2 are seen to follow rather smooth distributions as shown in **Figures 12** and **13** respectively. k_{synmax} also seems to follow a nice and smooth distribution, but with a few *outliers* scattering here and there as shown in **Figure 14**.

Note that those variance or confidence intervals for $Y_{X/S}$, k_2 , k_{synmax} by GSA are smaller than what was obtained by local search or the conventional GBM.

K_{IB} on the other hand presents totally different picture (**Figure 15**), that is, the model is almost insensitive to the variation of this parameter. We can actually confirm this by running sensitivity analysis separately as summarised in **Table 4**.

An important finding here is that the local search was *trapped* by the local minimum, but it is not necessarily a robust estimate because of the spread as shown in **Figure 15**.

3.2. Case study 'B': catalytic synthesis of methanol from CO and H₂

A laboratory integral reactor with fixed bed catalyst was studied. Model parameter estimation was attempted by the conventional maximum likelihood approach as well as the newly

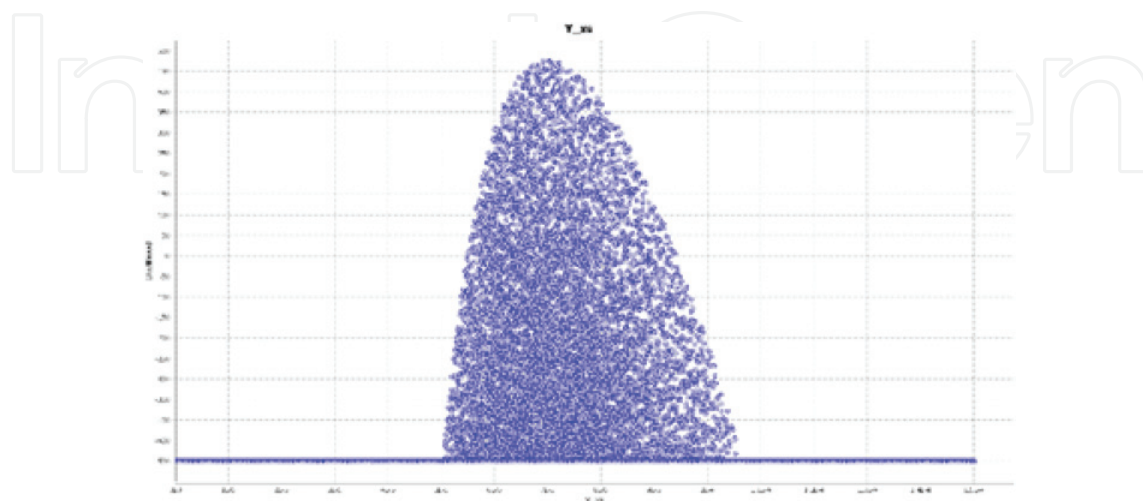


Figure 12. Marginal posterior ($Y_{X/S}$).

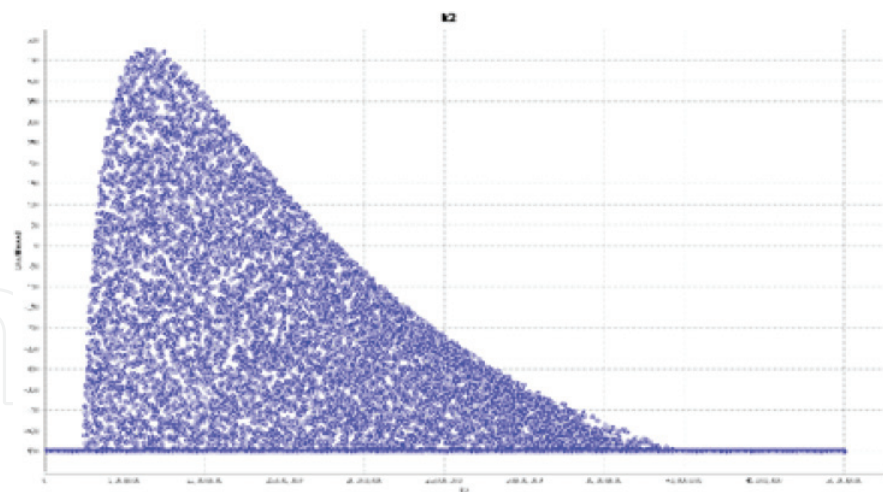


Figure 13. Marginal posterior (k_2).

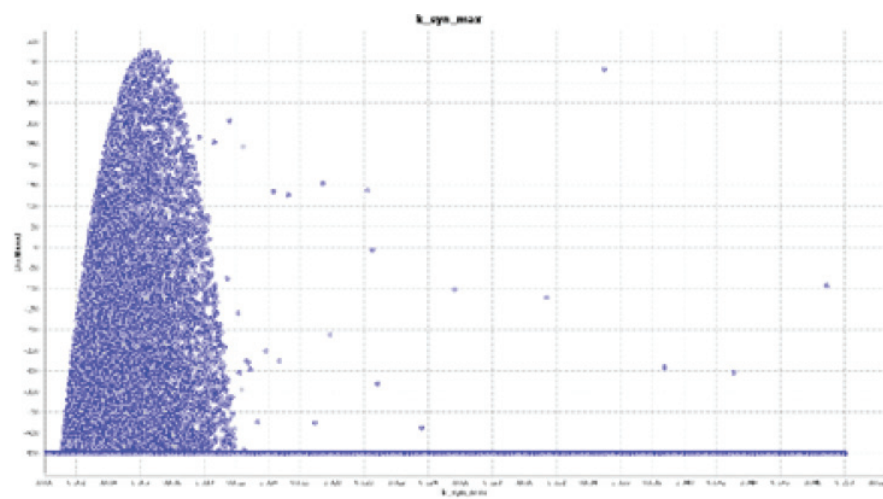
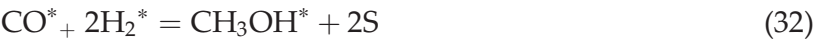


Figure 14. Marginal posterior (k_{synmax}).

proposed GSA methodology. Pseudo experiment data were prepared by introducing artificial measurement errors.

3.2.1. Problem statement

The following main reaction can be accomplished by several side and consecutive reactions (see [7]). A large number of kinetic expressions have been proposed in the literature [8]. We assumed the rate determining step (29) as the dominant reaction mechanism.



where * denotes component as adsorbed state; S is an adsorbed site.

A Langmuir-Hinshelwood type kinetics as described by Eq. 13 was assumed for the study.

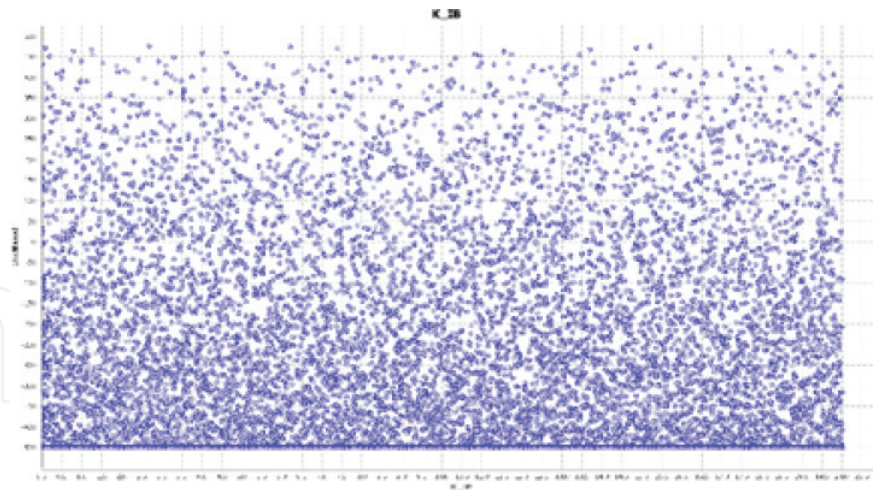


Figure 15. Marginal posterior (K_{IB}).

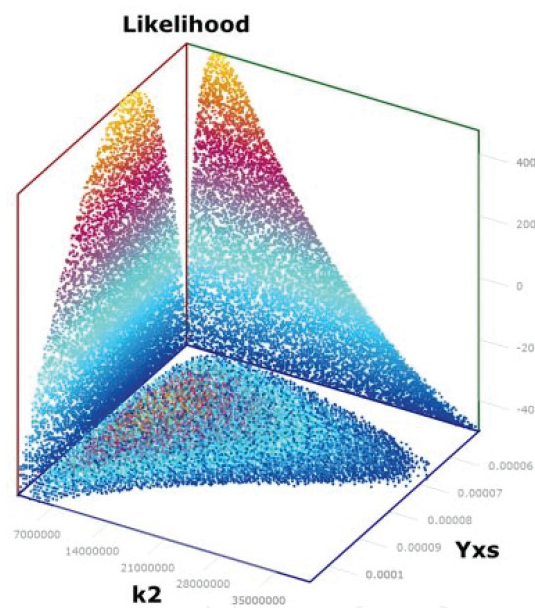


Figure 16. 2D marginal posterior ($Y_{X/S}$ vs. k_2).

$$r_{Methanol} = \frac{k(\phi_{CO}\phi_{H_2}^2 - \phi_{Methanol}/K_{eq})}{(1 + A\phi_{CO} + B\phi_{H_2} + C\phi_{Methanol})^2} \quad (33)$$

$$k = k_r \exp \left[\frac{-E_r}{RT} \right] \quad (34)$$

$$A = k_{CO} \exp \left[\frac{-E_{CO}}{RT} \right] \quad (35)$$

$$B = k_{H_2} \exp \left[\frac{-E_{H_2}}{RT} \right] \quad (36)$$

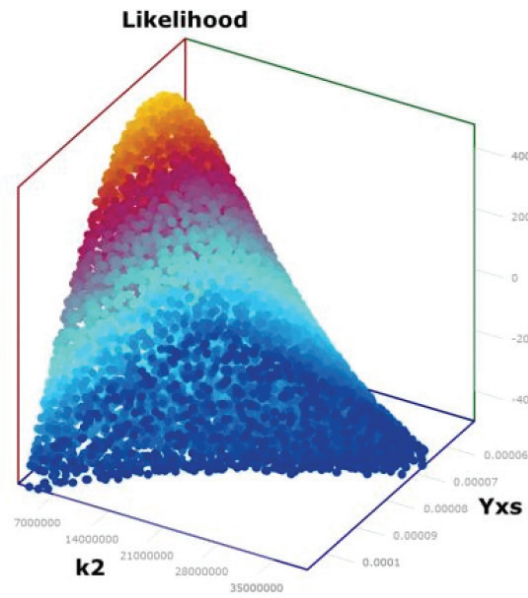


Figure 17. 3D posterior plot ($Y_{X/S}$ vs. k_2).

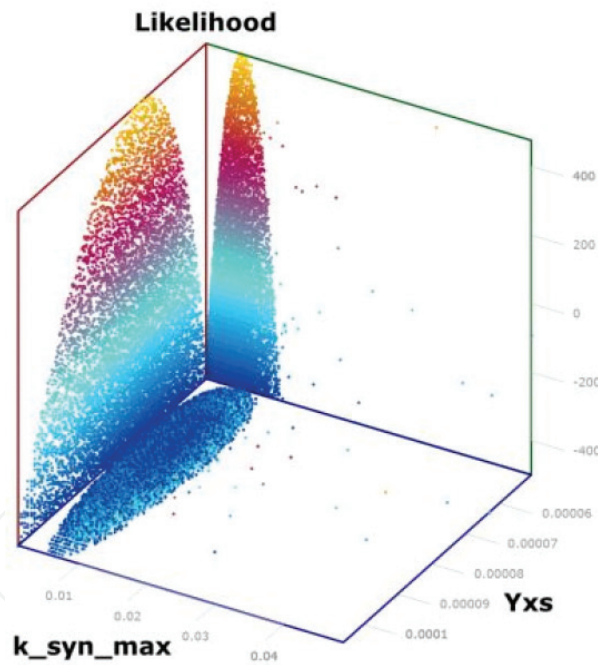


Figure 18. 2D marginal posterior ($Y_{X/S}$ vs. k_{synmax}).

$$C = k_{\text{Methanol}} \exp \left[\frac{-E_{\text{Methanol}}}{RT} \right] \quad (37)$$

where C_i , molar concentration of component i , (mol/m^3); E_j , activation energy of j th reaction, (J/mol); k_j , reaction pre-exponent factor of j th reaction, ($1/\text{sec}$); r_j , reaction rate of j th reaction,

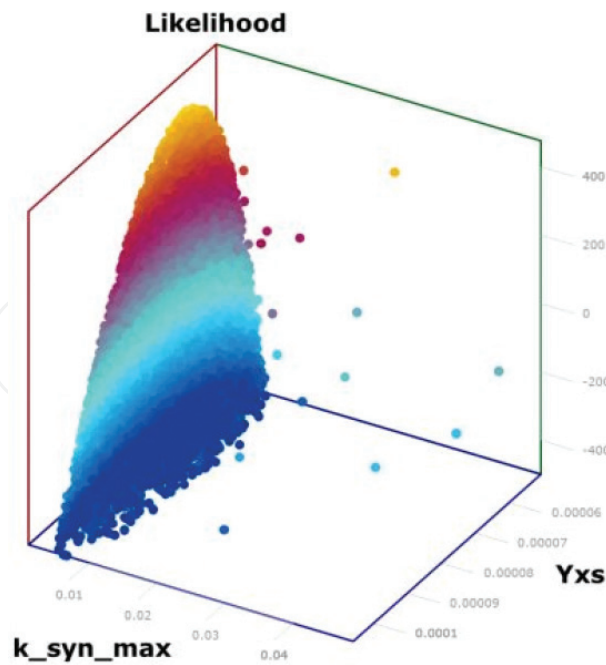


Figure 19. 3D posterior plot ($Y_{X/S}$ vs. k_{synmax}).

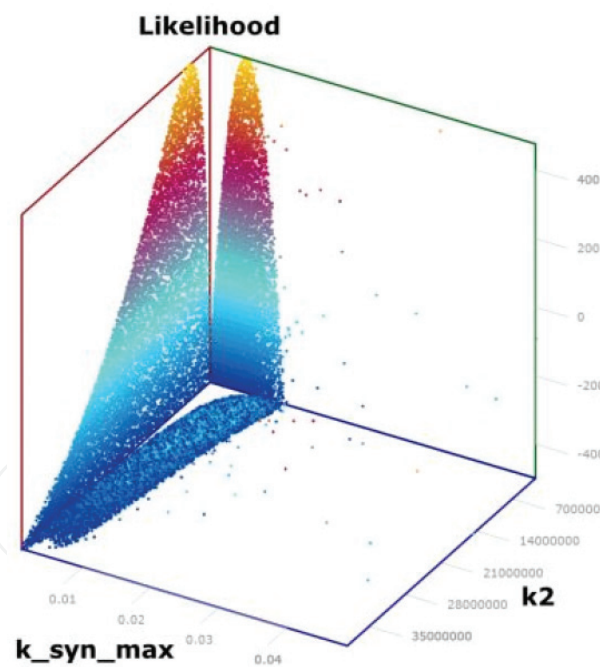


Figure 20. 2D marginal posterior (k_2 vs. k_{synmax}).

(mol/sec/m³); R , universal gas constant, (J/K/mol); t , time, (sec); and T , reaction (operating) temperature, (K).

The molar concentrations of the species are measured with Gaussian errors with mean at 0 and variance σ^2 .

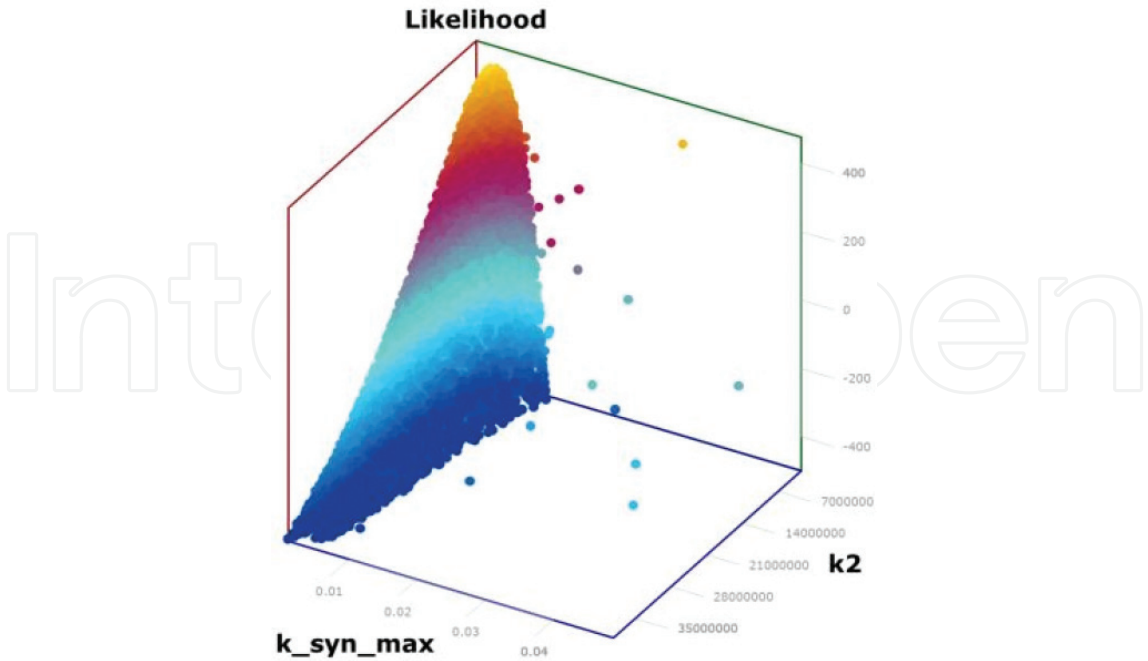


Figure 21. 3D posterior plot (k_2 vs. k_{synmax}).

	E_{CO}	E_{H_2}	$E_{Methanol}$	E_r	k_{CO}	k_{H_2}	$k_{Methanol}$	k_r
<i>Maximum likelihood estimators</i>								
Likelihood =	−3348							
MLE	6.72E + 04	5.13E + 04	7.16E + 04	1.22E + 05	7.05E + 05	3.41E + 03	1.97E + 06	1.01E + 09
Std. Dev.	1.65E + 04	1.69E + 04	1.61E + 04	4.92E + 04	2.95E + 06	1.46E + 04	8.07E + 06	1.26E + 10
<i>Empirical Bayes estimators (GSA-based)</i>								
Likelihood =	−3397							
MAP	6.56E + 04	4.86E + 04	7.29E + 04	1.21E + 05	8.35E + 05	3.59E + 03	4.47E + 06	4.58E + 09
Posterior Mean	6.55E + 04	4.86E + 04	7.29E + 04	1.21E + 05	8.35E + 05	3.60E + 03	4.47E + 06	4.58E + 09
Posterior Variance	6.48E + 02	4.49E + 02	6.93E + 02	1.67E + 03	1.21E + 04	1.20E + 02	8.04E + 04	2.26E + 07
<i>Factor sensitivity table</i>								
First order effect	0.004	−0.011	−0.031	0	−0.043	−0.019	0.02	−0.042
Total effect	0.607	0.256	0.69	0.885	0.179	0.078	0.271	0.209

Table 4. Case study ‘B’ estimated parameters.

$$\sigma^2 = \omega^2 z^2 \tag{38}$$

where, z is the measured data and ω ($=0.02$) is the standard deviation of the linear variance model. The reactor with a fixed catalyst bed was operated under three different temperatures (475, 500, 525, 550 K). Two different initial conditions were assumed, either being filled with the feed

gas compositions or an empty/inert state. There were eight sets of experiment data for the study. Feed rates were controlled as shown in **Figure 22**.

We can observe that the GSA or Bayesian approach correctly captures the process kinetics as described by Arrhenius equations that contains some interdependency among the activation energies and the pre-exponent factors.

3.2.2. Numerical study

Figures 23 and **24** show measurement data in a dynamic experiment (designated by the blue dots in the plots). Notice that the vertical bar attached to each dot denotes the measurement error that we assumed. The red dots are the simulation results of the final iteration for the MLE estimation by the GBM.

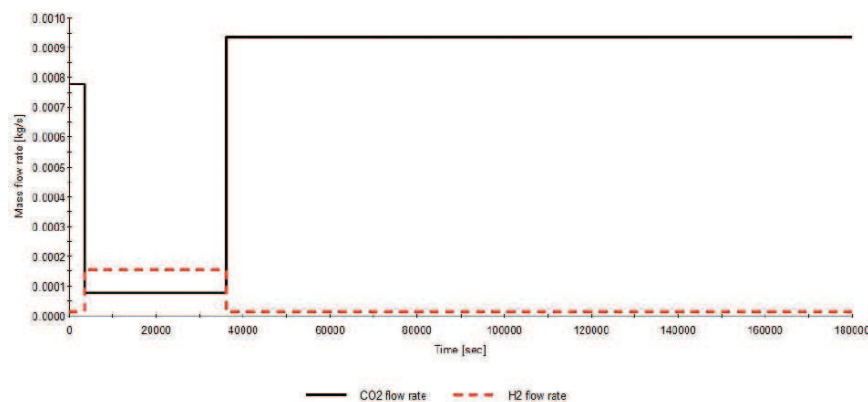


Figure 22. Feed rates profile.

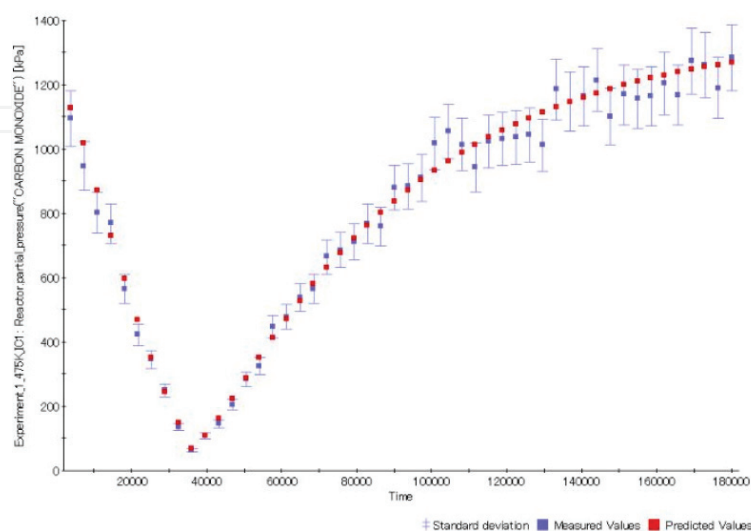


Figure 23. Trajectory of carbon monoxide yield.

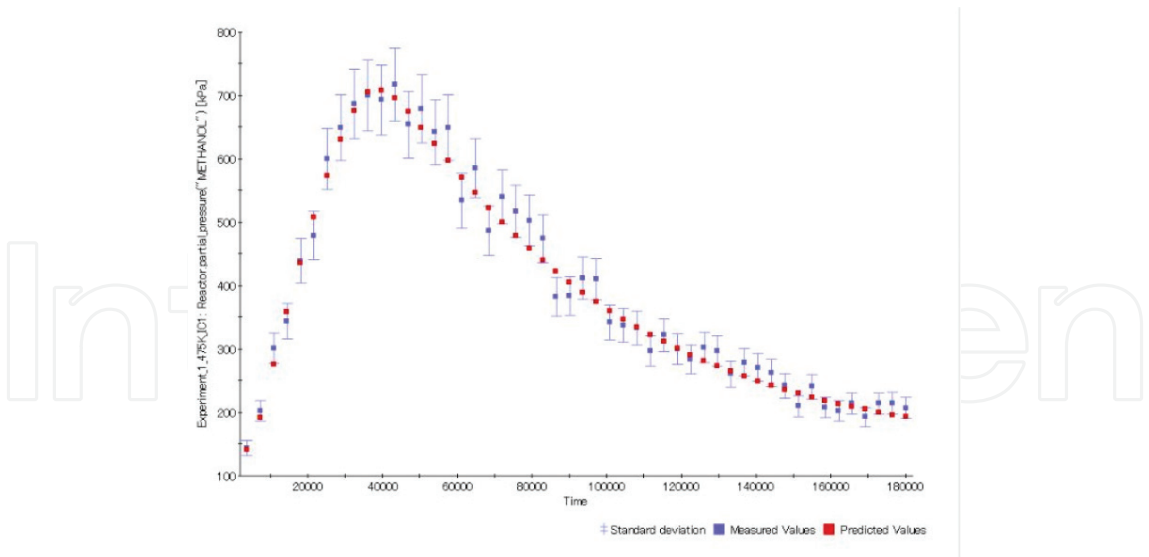


Figure 24. Trajectory of methanol yield.

The model parameters were estimated firstly by gPROMS[®], a *conventional* solution engine or GBM to maximise the likelihood function (minimising the log-likelihood). Confidence regions or intervals for the estimated parameters are evaluated based on Cramer-Rao lower bounds.

Second, posterior likelihood of the model parameters is computed by GSA. **Figures 25–30** are 3D plots of likelihood for a few pairs of parameters to be identified. The projection of those 3D points to 2D planes in the former plots are 2D marginal distributions as shown in **Figures 31–36**, respectively.

The parameters are correlated in one way or another. Nonetheless, the pair of parameters such as E_{CO} and E_r are still identifiable based on the posterior distribution of the likelihood function

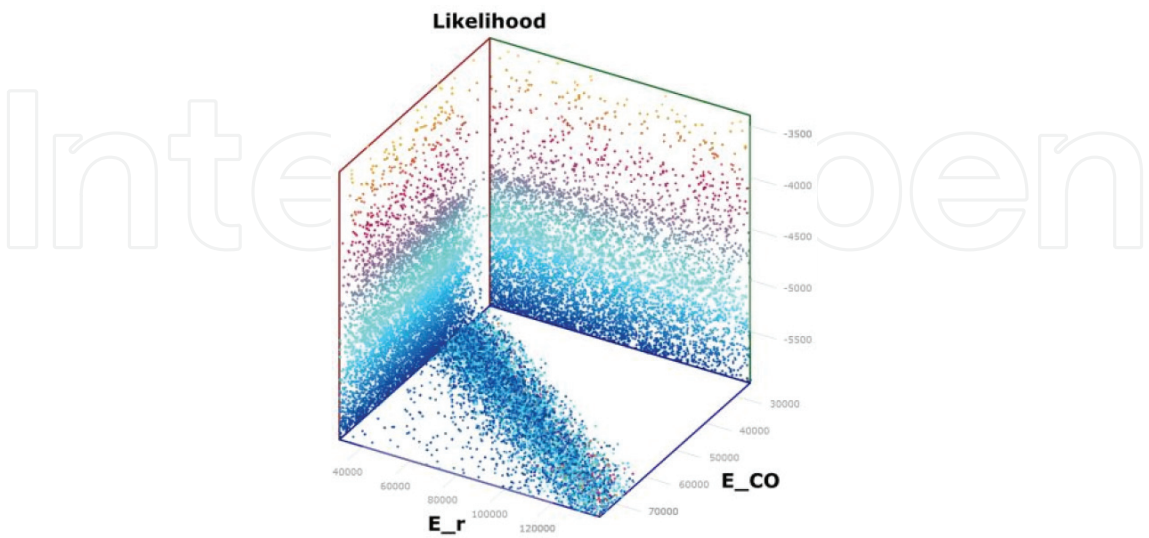


Figure 25. 3D posterior plot (E_{CO} vs. E_r).

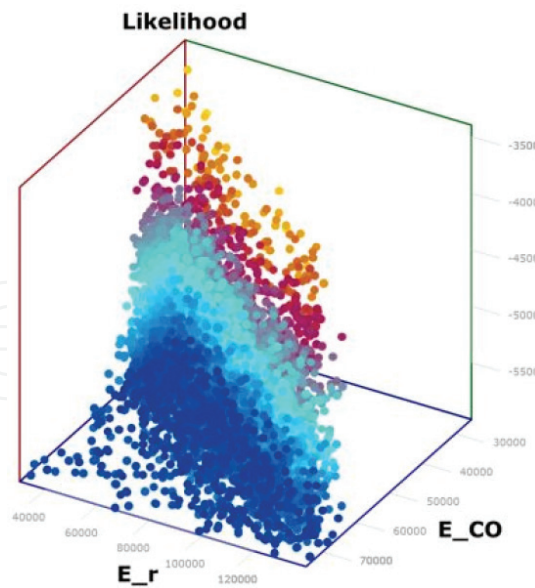


Figure 26. 3D posterior plot (E_{H_2} vs. E_r).

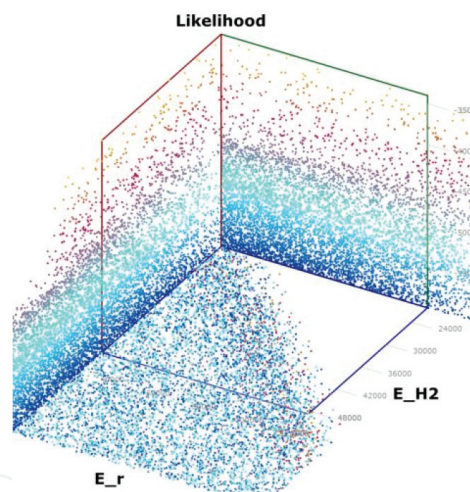


Figure 27. 3D posterior plot ($E_{Methanol}$ vs. E_r).

(**Figure 31**). The pair of E_{H_2} and E_r presents a little different profile, but they can also be estimated by taking MAP or averaging over the entire sampling space (**Figure 32**).

One noticeable finding was that the posterior variances of the GSA in this particular case were much smaller than those were estimated by local search or GBM (**Table 5**). Since the pre-exponent factors and the activation energies in the Arrhenius-type kinetics are highly correlated as can be seen in **Figures 34–36**, parameter estimation is difficult in its nature of the mathematical formulation.

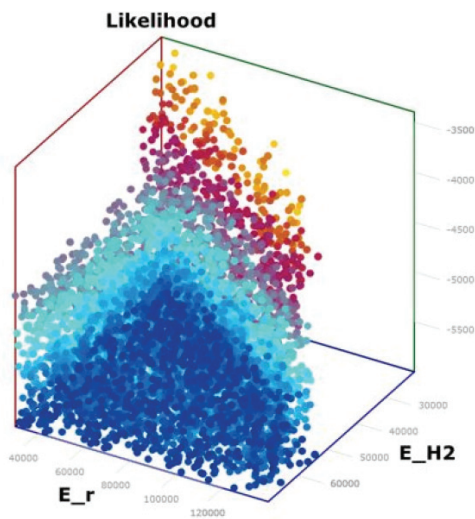


Figure 28. 3D posterior plot (E_{CO} vs. k_{CO}).

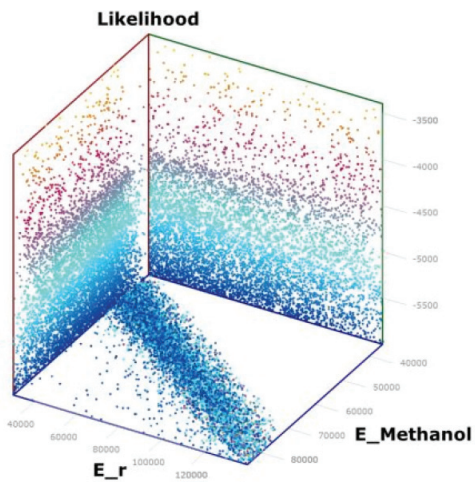


Figure 29. 3D posterior plot (E_{H2} vs. k_{H2}).

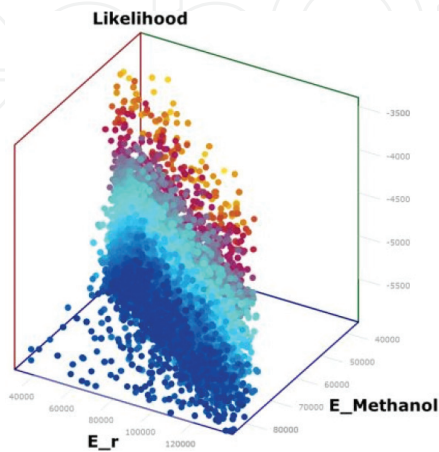


Figure 30. 3D posterior plot ($E_{Methanol}$ vs. $k_{Methanol}$).

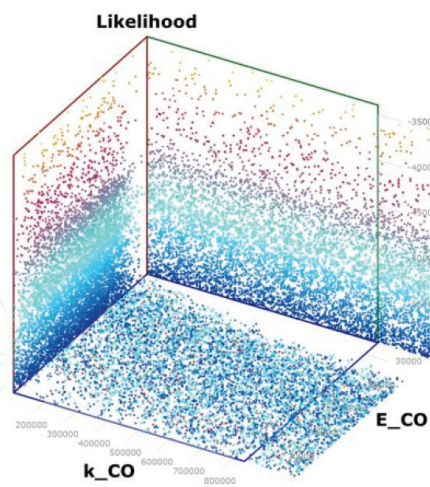


Figure 31. 2D marginal posterior (E_{CO} vs. E_r).

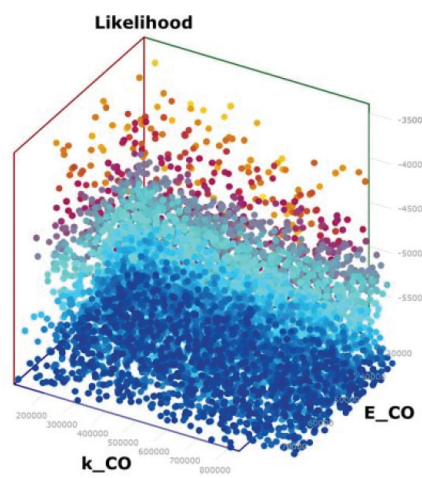


Figure 32. 2D marginal posterior (E_{H2} vs. E_r).

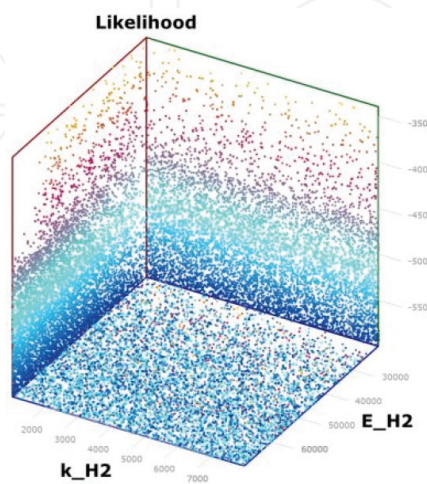


Figure 33. 2D marginal posterior ($E_{Methanol}$ vs. E_r).

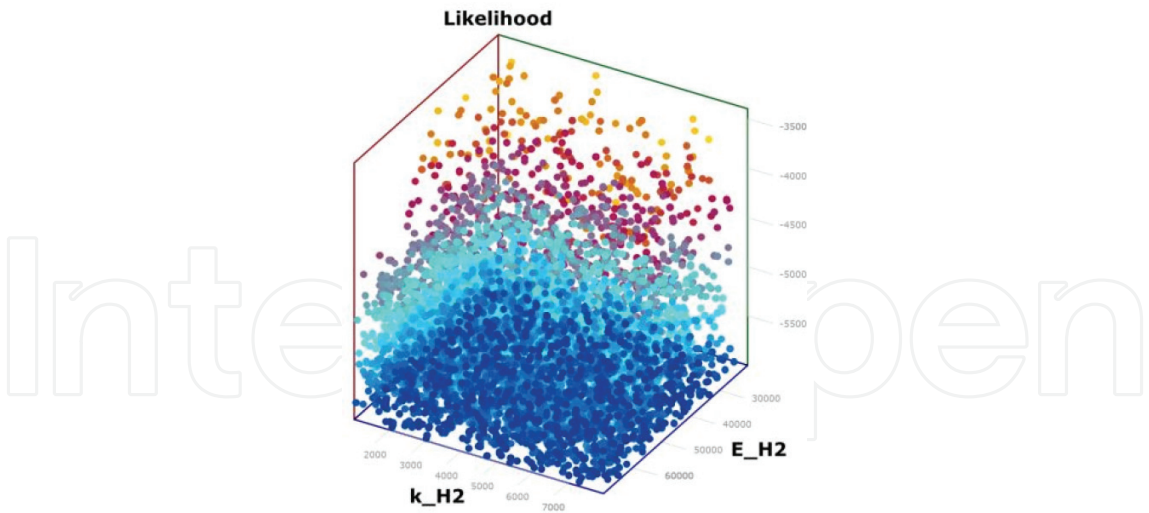


Figure 34. 2D marginal posterior (E_{CO} vs. k_{CO}).

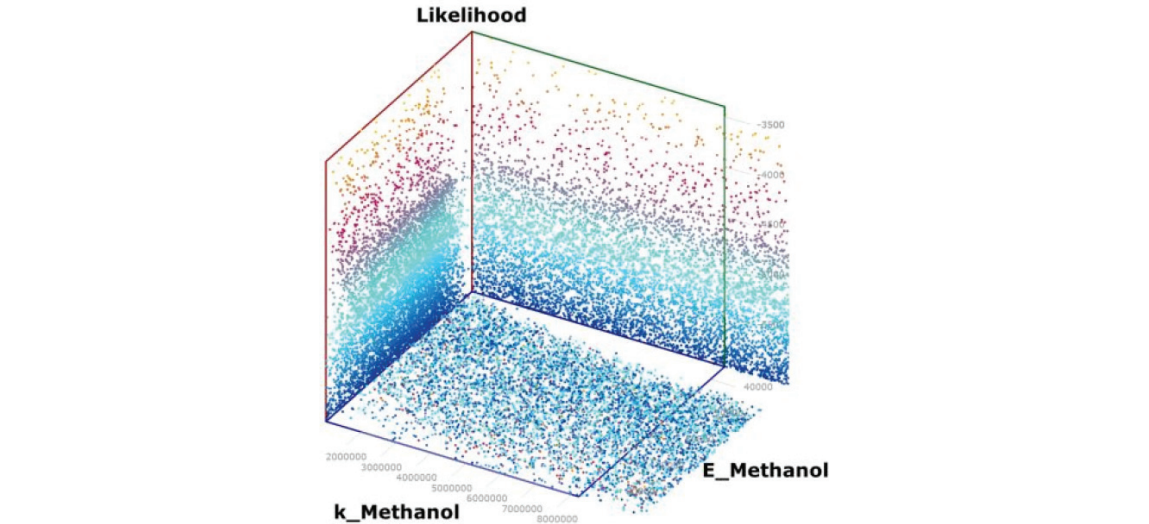


Figure 35. 2D marginal posterior (E_{H2} vs. k_{H2}).

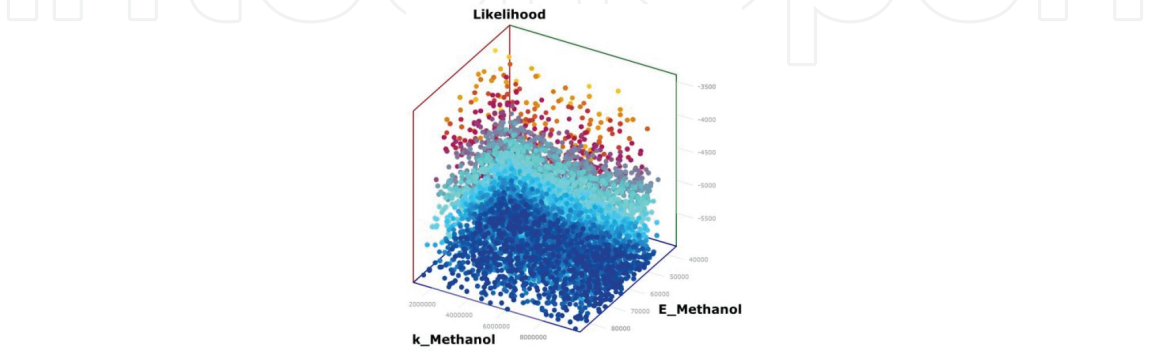


Figure 36. 2D marginal posterior ($E_{Methanol}$ vs. $k_{Methanol}$).

	μ	σ	μ	σ
	$0 < x_i < 10$		$0 < x_i < 1$	
x_1	111	60.5	1.11	0.65
x_2	30	13.4	3.0	1.34

Table 5. Elementary effects sensitivity analysis.

	S_i	$S_{i,T}$	S_i	$S_{i,T}$
	$0 < x_i < 10$		$0 < x_i < 1$	
x_1	0.92	1.02	0.03	0.12
x_2	-0.01	0.08	0.88	0.97

Table 6. Variance-based sensitivity analysis.

4. Conclusions

We applied GSA to parameter estimation for dynamic process models. Numerical case studies were performed for a couple of different processes with a set of simulated or artificial data. We demonstrated the advantage of the proposed methodology that is based on global search, over the conventional ones that are based on local search.

GSA calculates posterior distribution of the model parameters based on any measurement data available for the study. The estimated confidence region or intervals are in general narrower than those were calculated by the local search.

In case where some parameter is not affecting the system response significantly, GSA would still correctly evaluate the distribution or the probable spread of the model parameters. On the contrary, local search or GBM might find an optimal point for the same problem that can however be *elusive* in that the estimated variance around the solution is erroneously small due to lack of global information for the point estimate.

When a set of parameters are structurally correlated as in the case for the activation energy and pre-exponential coefficient, confidence region or confidence interval by the conventional local search methods can be very large thus smearing out the point estimate. Even under such circumstances, GSA can provide legitimate information about the accuracy of the estimated parameters. The posterior variance by GSA can be significantly smaller than Cramer-Rao lower bounds.

Therefore, we can conclude that GSA can provide more robust location parameters compared to the local search methods. But to be fair, because GBM can still find the optimal solution more precisely, it is advised to use both the methods complementarily: local search or GBM can initially be used to search for MLE. GSA can then provide a global view of the posterior probability of the point estimation to make sure the solution is stable and robust. GSA can also

be used for sensitivity study to make sure the estimated parameters or the variances of those are reasonable.

Since GSA is computation-intensive, parallel computation must be considered to practice such a method for a study of real-life problems. We made use of parallel computation on a single PC with multi-cores (employed four *workers*). GSA with 500,000 points sampling took CPU time of 9530 s (2.6 h) and 128,752 s (35.8 h) for the abovementioned example cases 'A' and 'B' respectively (performed on PC with Intel i7-5600 U @2.6GHz).

A. Appendix

A.1. Numerical implementation of Bayesian analysis: review of the incumbent method

Although we proposed a new way of calculating posterior probability by GSA as described in Section 2.2, it is worthwhile to review the incumbent method.

A.1.1. Markov chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a sampling method searching in a space. It is particularly important in Bayesian analysis for surveying a space with an arbitrary probability measure. When the posterior distribution is not obtained in a closed form, we need to rely on numerical sampling in the parametric space.

MCMC draws random samples from their marginal posterior distributions. With the introduction of prior distribution, we can compute the posterior distribution based on the Bayes' theorem.

A.1.2. Sampling from the posterior distribution

Metropolis-Hastings is an *accept-reject sampling*. Designing a good proposal distribution is the key to the success of such a sampling method. This concept is illustrated in the following pseudo steps.

- i. Propose a value $\theta^{(*)}$ for the parameter.
- ii. Compute the posterior probability ratio for the proposed value $\theta^{(*)}$ against the previous value $\theta^{(k-1)}$.
- iii. Accept the proposed value with the probability in the previous step (i.e. $\theta^{(k)} = \theta^{(*)}$). In case the proposal is not accepted, then retain the previous value (i.e. $\theta^{(k)} = \theta^{(k-1)}$).

where full conditionals are conjugate, Gibbs sampling can be used. A sufficient number of repetitions converge on the posterior distributions much more rapidly than by the Metropolis-Hastings sampling. When conditional distributions are not conjugate, we use Metropolis-Hastings.

A.1.3. Shortcomings of MCMC

Although MCMC has become a standard tool for Bayesian analysis, several problems have also become obvious particularly in the numerical implementation of the method [9]. While MCMC provides an almost automatic way of sampling the posterior distribution, it often converges too slowly or get stuck within one mode of a multi-modal parametric space. Moreover, it is oftentimes difficult to determine if a chain has reached stationarity or still hovering in a certain region of the parameter space.

The computation of the likelihood function can be complicated as in the physical process systems that are described by a set of mathematical equations, which can also be dynamic or time-variant rather than stationary or time-invariant. Therefore, the sampling must be efficient for the numerical implementation of MCMC to be applied to analysis of chemical, biological or physical systems that we encounter in real life.

A.2. Global sensitivity analysis: example calculations

A couple of example calculations of GSA sensitivity analysis are shown below (refer to Section 2.4).

A.2.1. Elementary effects sensitivity analysis

Applying elementary effects sensitivity analysis on equation $y = x_1^2 + 3x_2$ returns the results in Table 5.

A.2.2. Variance-based sensitivity analysis

Applying variance-based sensitivity analysis on equation $y = x_1^2 + 3x_2$ returns the results in Table 6. While variance-based values do not have an exact meaning, the most influential factors are correctly identified.

Author details

Shigeru Kashiwaya

Address all correspondence to: s.kashiwaya@psenterprise.com

Process Systems Enterprise Limited, Yokohama, Japan

References

- [1] Kashiwaya S. A Quasi-Monte Carlo approach to Bayesian parameter estimation for nonlinear dynamic process models. *Journal of Chemical Engineering of Japan*. 2013;46(7):467-479. DOI: 10.1252/jcej.12we202

- [2] Sobol' IM. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*. 1993;**1**(4):407-414
- [3] Sobol' IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematical Modelling and Computational Experiments*. 2001;**55**(1):271-280
- [4] Andrea S et al. *Global Sensitivity Analysis: The Primer*. London: John Wiley & Sons; 2008
- [5] Andrea S et al. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*. 2010;**181**(2):259-270
- [6] Andreas K et al. A benchmark for methods in reverse engineering and model discrimination: Problem formulation and solutions. *Genome Research*. 2004;**14**(9):1773-1785
- [7] Kuczynski M et al. Reaction kinetics for the synthesis of methanol from CO and H₂ on a copper catalyst. *Chemical Engineering and Processing*. 1987;**21**(4):179-191
- [8] Falbe J, editor. *New Synthesis with Carbon Monoxide*. Berlin: Springer; 1980. pp. 309-320
- [9] Guan Y et al. Markov chain Monte Carlo in small worlds. *Statistics and Computing*. 2006;**16**(2):193-202

IntechOpen